

На правах рукописи

Пермина Елизавета Алексеевна

**Изучение регулонов бактериального стресса
методами сравнительной геномики**

03.00.03 Молекулярная биология

А В Т О Р Е Ф Е Р А Т

диссертации на соискание ученой степени
кандидат биологических наук

– М о с к в а 2 0 0 6 –

Работа выполнена в лаборатории биоинформатики Государственного научно-исследовательского института генетики и селекции промышленных микроорганизмов ГосНИИ «Генетика».

Научные руководители:

кандидат физико-математических наук, доктор биологических наук
М. С. Гельфанд
кандидат физико-математических наук,
В. Ю. Макеев

Официальные оппоненты:

доктор физико-математических наук, профессор
В. Г. Туманян
кандидат биологических наук
И. В. Манухов

Ведущая организация: Институт Общей Генетики РАН (ИОГЕН)

Защита диссертации состоится «25» апреля 2006 г. в _____ ч. на заседании Диссертационного совета

С диссертацией можно ознакомиться в библиотеке ГосНИИГенетика

Автореферат разослан «23» марта 2006 г.

Ученый секретарь
Диссертационного совета,
кандидат биологических наук

В. И. Щербакова

Общая характеристика работы

Актуальность темы

Анализ регуляции экспрессии генов является важным разделом компьютерной генетики и системной биологии. Это связано с тем, что, с одной стороны, в настоящее время сравнительно легко осуществляется секвенирование геномов и эксперименты по анализу экспрессии на микрочипах, а с другой - использование этой информации для описания возможного фенотипа встречает трудности, обусловленные отставанием экспериментальной науки от темпов секвенирования.

Несмотря на этот значительный материал, многие детали регуляции не ясны до сих пор. В свете того, что экспериментальная проверка гипотез сложна и требует значительного количества времени и ресурсов, представляется целесообразным разработать и использовать методы, которые позволяют перенести часть исследований в область компьютерной биологии. Подобные исследования дают возможность определять наиболее перспективные направления для дальнейшей экспериментальной проверки, а в некоторых случаях делать утверждения, практически равносильные результатам биологического эксперимента.

Изучение регуляции ответа на стресс тем более важно, что позволит координировать микробиологические исследования в области поиска новых антибактериальных агентов и техник дезинфекции. Изменчивость и лабильность бактериального генома позволяет патогенным организмам быстро приспосабливаться к уже разработанным антибиотикам и техникам дезинфекции, и исследования в этой области сохраняют актуальность в течение всего периода развития микробиологии и медицины. В частности, рассмотренные в данной работе регуляторные каскады общего стресса в группе *Bacillus* показывают связь нескольких бактериальных стрессовых регулонов с множественной лекарственной устойчивостью.

Быстрое развитие методов индустриальной биоинженерии также требует всестороннего изучения изменения клеточных процессов в необычных для клетки условиях биотехнологического процесса. Новое направление в этой области ставит своей задачей разработать методики восстановления окружающей среды от загрязнений ионами тяжелых металлов и других биологически опасных веществ. Изучение микроорганизмов, способных осаждать соли тяжелых металлов из раствора, является одной из ключевых задач в этой области.

Цель и задачи исследования

Диссертация состоит из четырех глав, посвященных регуляции SOS-ответа, теплового шока, общего стресса и устойчивости к ионам тяжелых металлов.

Целью исследования, представленного в первой и второй главах, является применение современных методов сравнительной геномики к исследованию ответа на повреждение ДНК и тепловой шок у бактерий. При этом решались следующие задачи:

- Описание структуры всех генетических локусов, содержащих гены ответа на повреждение ДНК и тепловой шок в группе гамма-протеобактерий и Грам-положительных бактерий.
- Поиск в геномах бактерий из группы гамма-протеобактерий и фирмикут новых генов, относящихся к системам ответа на повреждение ДНК и тепловой шок.
- Анализ пересечений различных регулонов теплового шока, изучение регуляторных каскадов.

Целью исследования, представленного в третьей главе, является изучение регулона общего стресса (SigB) в группе *Bacillus/Staphylococcus/Listeria*:

- Описание структуры всех генетических локусов, содержащих гены, входящие в систему общего стресса у шести геномов группы *Bacillus/Staphylococcus/Listeria* (*B. subtilis*, *B. halodurans*, *B. anthracis*, *S. aureus*, *O. ihiensis*, *L. monocytogenes*). Были выбраны полные геномы, в которых присутствует ген регулятора.
- Сравнение данных, полученных с помощью методов сравнительной геномики, с данными, полученными с помощью метода микрочипов, взятыми из литературы.

Целью исследования, представленного в четвертой главе, является изучение регулонов ответа на присутствие в среде ионов тяжелых металлов:

- Описание структуры всех генетических локусов, содержащих гены, входящих в системы устойчивости к ионам тяжелых металлов, последовательности которых доступны на момент проведения исследования.
- Сравнительный анализ эволюции транскрипционных регуляторов и регуляторных сигналов систем устойчивости к ионам тяжелых металлов.

Научная новизна и практическое значение

В работе впервые получены следующие результаты:

- Впервые систематически описана структура всех локусов, содержащих гены регуляторных систем SOS-ответа, теплового шока, общего стресса и систем устойчивости к ионам тяжелых металлов.
- Для бета-протеобактерий описана система взаимодействия регулонов теплового шока *RpoH* и *HrcA* и регуляторный каскад.
- Для гамма- и бета-протеобактерий предсказано вхождение гена *ybbN* в регулон теплового шока.
- Для Грам-положительных бактерий предположен ряд новых сайтов связывания регулятора *HrcA*.
- Проведено сравнение результатов сравнительно-геномного анализа и опубликованных данных по анализу экспрессии на микрочипах.
- Предсказана специфичность ряда регуляторов и транспортеров, относящихся к системе устойчивости к ионам тяжелых металлов.

Всё вышесказанное составляет практическую значимость работы и является поводом для экспериментальной проверки.

Апробация работы

Результаты, изложенные в диссертации, докладывались и обсуждались на международных научных конференциях BGRS (International Conference on Bioinformatics of Genome Regulation and Structure, Новосибирск 2002, 2004) и MCCMB (Moscow Conference in Computational Molecular Biology, Москва 2003, 2005), а также на научных семинарах в лаборатории Биоинформатики ФГУП ГосНИИгенетика, Москва, Россия; Ecole Polytechnique, Palaiseau, Франция; National Center of Biological Information, Bethesda, США; Instituto de Biologia Molecular, Rosario, Аргентина. Они опубликованы в шести тезисах упомянутых конференций.

Объем и структура диссертации

Диссертация изложена на страницах машинописного текста и состоит из введения, обзора литературы, описания материалов и методов, результатов (четыре главы), выводов и списка литературы, содержащего наименований. Работа содержит рисунков и таблиц.

Глава I посвящена сравнительному анализу систем устойчивости к агентам, повреждающим ДНК у гамма-протеобактерий и Грам-положительных бактерий. Описано консервативное ядро регулона SOS-ответа в группе гамма-протеобактерий, впервые описаны таксон-

специфические черты SOS-регулона внутри каждого рода, предсказано участие в SOS-ответе двух новых генов. Описана система ответа на повреждение ДНК у Грам-положительных бактерий. Дана характеристика распространения системы UmuDC.

Глава II посвящена исследованию генетических аспектов системы теплового стресса в гамма-протеобактериях и Грам-положительных бактериях. Произведена классификация изученных регуляторных механизмов и построение схемы взаимодействия различных регулонов теплового шока.

Глава III содержит исследование системы общего стресса в группе *Bacillus/Staphylococcus/Listeria*. Описана структура регулона в шести геномах, выделена консервативная часть регулона и проведено сравнение данных, полученных при помощи метода микрочипов, и результатов сравнительно-геномного анализа.

В **Главе IV** проведено описание и анализ систем устойчивости к ионам тяжелых металлов. Описана схема определения специфичности регуляторов семейства MerR/HMR и транспортеров ионов тяжелых металлов, предсказано участие в ответе на повышение концентрации ионов Cu[I] цитохромов c553 в геномах *Vibrio cholerae*, *V. vulnificus* и *V. parahaemolyticus*.

Методы, применявшиеся в работе

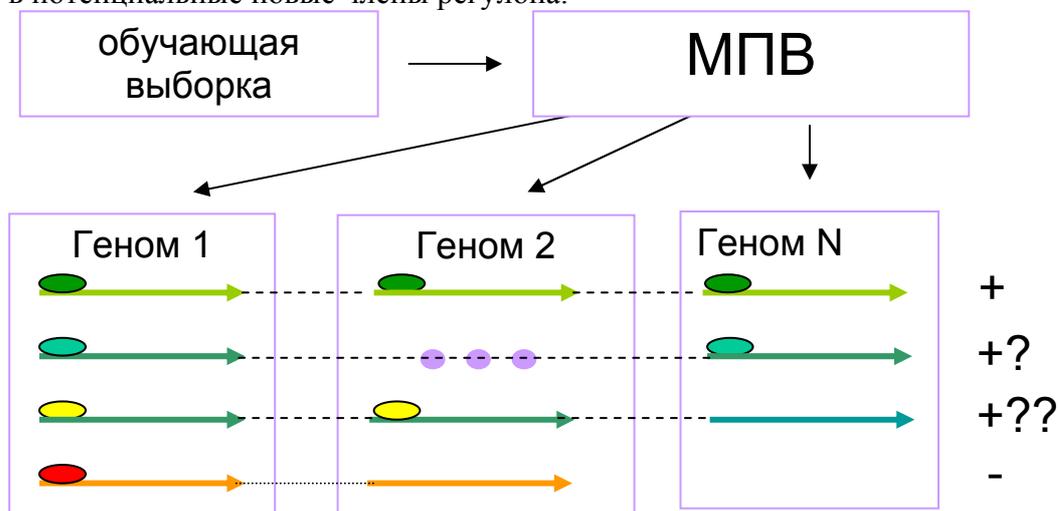
Метод сравнительной геномики

Для сортировки первичных данных, полученных с помощью матриц позиционных весов (рабочей выборки), и разделения их на перепредсказанные и потенциальные сайты, был использован метод сравнительной геномики (филогенетического футпринтинга). Рабочие выборки из нескольких родственных геномов сравнивались с целью найти группу ортологичных генов с консервативным потенциальным сайтом в 5'-некодирующей области (Рис. 1). Если потенциальные сайты встречались в трех или более случаях, ген определялся как предсказанный участник регулона. При анализе также учитывалась известная или предполагаемая функция гена, положение потенциального сайта и степень консервативности некодирующей области.

Базовым геномом становился, как правило, тот, из которого были взяты сайты для построения матрицы позиционных весов (МПВ), т.е. наиболее экспериментально изученный. Естественным ограничением такого метода является неучет случаев, когда ген утратил регуляцию именно в базовом геноме. Для компенсации этого применялся метод попарных сравнений. В этом случае сравнивались рабочие выборки всех возможных пар геномов в изучаемой группе. Этим методом определялась таксон-специфическая регуляция.

Рисунок 1. Схема сравнительно-геномного анализа

Стрелками обозначены гены, пунктирной линией – ортологичность, овалами обозначены потенциальные сайты. Плюсы стоят напротив генов, которые признаются потенциальными членами регулона, минус означает, что ген не попадает в потенциальные новые члены регулона.



Содержание работы

Глава I. SOS-ответ у гамма-протеобактерий и Грам-положительных бактерий

Система ответа на повреждение ДНК (система SOS-ответа) может быть индуцирована UV-излучением или некоторыми химическими агентами (митомycin). Внутриклеточным событием, включающим регуляторный каскад SOS-ответа, является появление одноцепочечной ДНК, с которой реагирует белок RecA с образованием нуклеопротеиновых филаментов. Активация белка RecA, в свою очередь, вызывает автопротеолиз белка-репрессора SOS-ответа LexA (Walker, 1984; Witkin, 1976; Rupp, 1999).

В SOS-ответе участвуют белки рекомбинационной и экзцизионной репарации (продукты оперонов *ruvAB* и *uvrABC*), гипермутирующая полимеразы UmuDC, белки, регулирующие клеточное деление (*SulA*) (Rupp, 1999; Sutton, 2000).

Был проведен анализ распространения и структуры SOS-регулонов в гамма-протеобактериях и Грам-положительных бактериях.

Исследование SOS-регулона методами сравнительной геномики

МПВ (матрица позиционных весов) строилась по экспериментально подтвержденным сайтам перед генами *lexA*, *recA*, *recN* и *ruvAB* *E. coli* (Rupp, 1999). Методами сравнительной геномики было выявлено

консервативное ядро регулона в гамма-протеобактериях (гены *lexA*, *recA*, *recN*) и показаны возможные составы регулона в родственных *E. coli* геномах, в частности, в *Salmonella typhimurium*, *Yersinia pestis*, *Pseudomonas aeruginosae*, *Haemophilus spp.* и *Xylella spp.* Было описано распространение оперона *umuDC* (*mucAB*) в геномах гамма-протеобактерий и феномен сопряжения оперона *umuDC* с плазмидами. В группе гамма-протеобактерий было проведено исследование возможной таксон-специфической регуляции для родов *Vibrio* и *Pseudomonas* и групп *Enterobacteriaceae* и *Pasteurellaceae* и выявлены характерные особенности организации SOS-регулона в этих таксономических группах.

В группе Грам-положительных бактерий также были охарактеризованы потенциальные регулоны ответа на повреждение ДНК. Так как регуляторная последовательность, с которой связывается регулятор DinR (ортолог LexA в *B. subtilis*), отличается от SOS-бокса *E. coli*, для Грам-положительных бактерий была построена своя матрица позиционных весов на основе выборки экспериментальных сайтов из *B. subtilis*.

Характеристика особенностей распространения и регуляции оперона umuDC, кодирующего Pol V

В геноме *B. subtilis* наблюдалось сопряжение оперона гомологичного *umuDC* с профаговым элементом, в то время как в *Enterococcus faecalis* гомолог *umuDC* располагается в одном локусе с генами, характерными для траспозонов. Из этих наблюдений можно сделать вывод о частой локализации оперонов *umuDC* в мобильных генетических элементах. Во всех случаях, когда локус содержал в себе оба гена *umuD* и *umuC*, локус находился под потенциальной регуляцией репрессора SOS-ответа, как в гамма-протеобактериях, так и в Грам-положительных бактериях. В случаях, когда локус не содержал гена *umuD*, регуляция могла быть утеряна. Выдвинуто предположение, что, поскольку полный оперон кодирует мутагенную полимеразу, он должен жестко репрессироваться в нормальных условиях во избежание повреждения генома. Гомолог *umuC* в отсутствие *umuD*, возможно, не обладает мутагенной способностью и его репрессия факультативна.

Глава II. Регулоны теплового шока у протеобактерий и Грам-положительных бактерий

Были рассмотрены регулоны HrcA, RpoH, RpoE, CtsR, SigB. Из них HrcA и CtsR являются негативными регуляторами транскрипции, содержащими ДНК-связывающий домен вида спираль-поворот-спираль, а RpoH, RpoE и SigB - активаторами транскрипции (альтернативными сигма-факторами).

HrcA характерен тем, что связывается с высоко консервативным палиндромом со структурой TTAGCACTC-(9)-GAGTGCTAA (Shumann et al., 1994). В общем случае в геноме находится один-два сайта связывания. В большинстве случаев сайт связывания HrcA находится перед опероном *groESL*, кодирующим один из главных шаперонов, задействованных в тепловом стрессе, и, иногда, перед геном *hrcA* и опероном *dnaKJ*. Регулон HrcA широко распространен среди бактерий, однако его представители практически отсутствуют в группе гамма-протеобактерий. Впервые он был описан в *B. subtilis* (Shumann et al. 1997), и его ортологи найдены в бета- и альфа-протеобактериях и широко распространены в группе Грам-положительных бактерий (Narberhaus, 1999).

RpoH (σ^{32}) представлен в большом количестве геномов протеобактерий. Регулятор σ^{32} - альтернативный сигма фактор, узнающий промотор с консенсусом CTTGAAA-(12)-CCCCAT (Gross et al. 1985, Gross, 1996). Регулон RpoH включает в себя гены, кодирующие основные белки теплового стресса - шапероны (GrpE, GroESL, DnaKJ, HtpG) и протеазы (Clp, Lon). В *E. coli* также описан второй регулон теплового шока - RpoE (σ^{24}). Промотор, узнаваемый σ^{24} , найден перед геном *rpoH* (Sihavy et al., 1997; Georgopoulos et. al, 1996; Yura et. al, 2000)

Ортологи гена *ctsR* встречены в группе Грам-положительных бактерий вплоть до группы Clostridia. Сигнал представляет собой тандемный повтор с шагом 7 нуклеотидов GTCAAAA...GTCAAAA (Msadek et. al, 1998, Msadek et al., 1999, Msadek et al., 2000).

Ниже приведены результаты анализа генетических аспектов реакции теплового шока. Для описания локусов защиты от повышенной температуры и их регуляции в геномах протеобактерий и Грам-положительных бактерий, для которых отсутствуют экспериментальные данные, были применены методы сравнительной геномики.

Анализ распространения и структуры регулонов теплового шока

Были описаны потенциальные регулоны HrcA в 17 геномах. Помимо Грам-положительных геномов из родов *Bacillus*, *Staphylococcus*, *Enterococcus*, *Clostridium*, *Lactobacillus*, *Lactococcus*, *Mycoplasma* и *Ureaplasma*, были рассмотрены геномы альфа-, бета-, гамма- и эпсилон-протеобактерий, в которых также были найдены ортологи HrcA (из группы гамма-протеобактерий регулон HrcA найден в *Xylella spp.* и *Thiobacillus spp.*) По литературным данным (Narberhaus, 1999) была построена МПВ для поиска потенциальных сайтов связывания регулятора. Ввиду того, что сайт связывания HrcA обладает чрезвычайно высокой консервативностью, порог для МПВ был выбран таким образом, чтобы допускать не более 4 замен. Были описаны потенциальные сайты связывания в 20 геномах. В большинстве случаев регулон HrcA в геномах Грам-положительных бактерий представлен оперонами *groESL* и *hrcA-grpE-dnaKJ*. В геномах *Clostridium* и *Mycoplasma* были впервые обнаружены потенциальные сайты связывания перед генами протеаз.

Ортологи гена *groH* из *E. coli* прослеживаются до альфа-протеобактерий. При анализе распространения регулона RpoH было выявлено сохранение сигнала вплоть до бета-протеобактерий. Было определено консервативное ядро регулона для гамма- и бета-протеобактерий, и впервые предположена возможность участия в системе теплового стресса гена *ybbN* *E. coli* (b0492), кодирующего тиоредоксин, на основании консервативности потенциального сайта перед ортологами гена, как в гамма-, так и в части бета-протеобактерий.

Группа бета-протеобактерий представляет особый интерес с точки зрения изучения регуляции теплового стресса, так как в геномах бактерий из этого таксона встречаются как ортологи HrcA, так и регулон RpoH. Нами впервые был описан регуляторный каскад HrcA->RpoH в геномах шесть бактерий (*A. xylooxidans*, *B. pertussis*, *B. parapertussis*, *B. pseudomallei*, *M. flagellatus*, *Methylovorus sp.*, *N. europaea* и *C. metallidurans*). В геномах трех из них впервые был обнаружен потенциальный промотор RpoH перед геном *hrcA* (Таблица 2), что приводит к образованию обратной связи HrcA<->RpoH.

Для CtsR консервативное ядро регулона описано в 12 геномах. МПВ создавалась на основе экспериментально показанных сайтов перед генами *clpE*, *clpP* и *clpC* из *B. subtilis* (Msadek et al. 1999). Рабочая выборка в *B. subtilis* составила 41 сайт у 37 генов. Методом попарных сравнений были выделены потенциальные сайты связывания CtsR в геномах *B. subtilis*, *Staphylococcus aureus*, *Streptococcus pneumoniae*, *S. pyogenes*, *Lactococcus lactis*, *Listeria monocytogenes*, *Enterococcus faecalis*, *Clostridia acetobutylicus*, *C. botulinus*, *C. perfringens*, *C. difficile*. Консервативное ядро регулона составляют протеазы Clp, и их регуляция сохраняется во всех рассмотренных геномах. В геномах *Enterococcus faecalis*, *Lactococcus lactis* и *Listeria monocytogenes* также обнаружен

потенциальный сайт перед геном *hrcA*, что позволяет предположить наличие регуляторного каскада CtsR->HrcA.

Таблица 1. Консервативное ядро регулона RpoH в гамма-протеобактериях

	ST(c)	YP(c)	VC(c)	PA(c)	XF(c)	BP	NE	BC
<i>dnaKJ</i>	+	+	+	+	(+)	+	+	+
<i>grpE</i>	+	+	+	-	+	+	+	+
<i>groESL</i>	+	+	+	+	+	+	+	+
<i>htpG</i>	+	+	+	+	+	+	+	+
<i>ftsJ</i>	+	+	+	0	-	-	0	-
<i>lon</i>	+	+	+	+	+	+	-	-
<i>clpB</i>	+	+	+	+	+	+	+	+
<i>clpP</i>	+	+	+	-	+	-	-	-
<i>clpX</i>	-	-	(+)	+	-	-	-	(+)
<i>hflB</i>	(+)	(+)	(+)	+	-	-	-	+*
<i>hslU</i>	-	+	-	(+)	(+)	-	+	+
<i>hslV</i>	0	+	+	+	+	+	-	+
<i>hslT</i>	+	+	+	0	0	0	0	0
<i>b0492</i>	+	+	+	-	+	0	0	0

Обозначения:

"+" – присутствие потенциального σ^{32} промотора;

"-" - отсутствие потенциального σ^{32} промотора;

"0" - отсутствие гена в секвенированной части генома;

"(+)" - ген расположен в потенциально регулируемом опероне.

* BC содержит две копии гена *hflB*.

Обозначения геномов ST - *S. typhi*, YP - *Y. pestis*, VC - *V. cholerae*, PA - *P. aeruginosa*, XF - *X. fastidiosa*, BP - *B. pertussis*, NE - *N. europaea*, BC - *Burkholderia cepacia*.

"(c)" - полный геном на момент проведения исследования.

Таблица 2. Потенциальные сайты связывания в 5'-некодирующих областях генов *rpoH* и *hrcA* в геномах бета-протеобактерий.

геном	ген	позиция	сайт связывания HrcA
<i>Achromobacter xylosoxidans</i>	<i>rpoH</i>	-38	TTAGCACTC-(9)-GAGTGCTAA
<i>Ralstonia eutropha</i>	<i>rpoH</i>	-97	TTAGCACTC-(9)-GAGTGCTAA
<i>Ralstonia solanacearum</i>	<i>rpoH</i>	-69	TTAGCACTC-(9)-GAGTGCTAA
<i>Bordetella pertussis</i>	<i>rpoH</i>	-73	TTAGCACTC-(9)-GAGTGCTAA
<i>Bordetella parapertussis</i>	<i>rpoH</i>	-92	TTAGCACTC-(9)-GAGTGCTAA
<i>Burkholderia pseudomallei</i>	<i>rpoH</i>	-87	TTAGCACTC-(9)-GAGTGCTAA
<i>Nitrosomonas europaea</i> *	<i>rpoH</i>	-84	TTAGCACTC-(9)-GAGTGCTAg
<i>Methylobacillus flagellatus</i>	<i>rpoH</i>	-56	cTAGCACaC-(9)-GAGTGCTAg
<i>Methylovorus sp. SS1</i>	<i>rpoH</i>	?	TTAGCACTC-(9)-GAGTGCTAG
Сайт связывания RpoH			
<i>Bordetella pertussis</i>	<i>hrcA</i>	-27	gTTGAAA-(15)-gCtCAT
<i>Bordetella parapertussis</i>	<i>hrcA</i>	-25	gTTGAAA-(15)-gCtCAT
<i>Methylobacillus flagellatus</i>	<i>hrcA</i>	-39	aTTGAAa-(14)-cCtCAT

Глава III. Регулон SigB: Общий стресс в группе *Bacillus/Staphylococcus/Listeria*.

Регуляторный белок SigB является альтернативным сигма фактором, распознающим промотор с консенсусной последовательностью

AGTTTAC-(14)-GGGTAT (Price, 2000; <http://dbtbs.hgc.jp>). Регулон SigB в *B. subtilis* достаточно обширен (по разным оценкам, от 75 до 200 генов; Price, 2000) и отвечает за устойчивость к повышению температуры, понижению pH, воздействию этанола и осмотический шок. В регулон SigB входят протеазы теплового шока (*clp*), которые, кроме того, входят у *B. subtilis* в регулон CtsR. В литературе была описана регуляция самого гена *ctsR* при помощи SigB (Kruger et al., 1996) Кроме того, в общем стрессе принимают участие белки-осмопротекторы (OpuE), каталазы (KatX, KatB), возможно участие белков множественной лекарственной устойчивости (Bmr) (Petersohn et al., 1999).

Исследование регулона общего стресса методом сравнительной генетики

В работе изучались консервативное ядро регулона и структура потенциальных регулонов в геномах группы *Bacillus/Staphylococcus/Listeria* и степень соответствия данных сравнительной геномики данным, полученным экспериментальными методами массового анализа уровня экспрессии генов.

По экспериментально показанным сайтам связывания SigB (<http://dbtbs.hgc.jp>) была построена МПВ. Рабочая выборка (число генов, содержащих в 5'-некодирующей области потенциальные сайты до фильтрации методом сравнительной геномики) для *B. subtilis* составляла 456 генов. Так как ортологи гена *sigB* на момент исследования были обнаружены в *B. anthracis*, *B. halodurans*, *S. aureus*, *O. ihiensis* и *L. monocytogenes*, и сигнал связывания в них консервативен, был проведен сравнительно-геномный анализ. Было выяснено, что консервативное ядро регулона сравнительно невелико (11 генов), и в рассмотренных геномах было найдено от 15 (*L. monocytogenes*) до 49 (*O. ihiensis*) генов, потенциальные сайты перед которыми встречались еще хоть в одном геноме (Таблица 3). Всего 11 генов обладают потенциальными сайтами в четырех или более геномах из семи рассмотренных.

Сравнение результатов, полученных методом сравнительной геномики с данными анализа экспрессии на микрочипах

Общий стресс неоднократно становился предметом исследования, и в 2001 году сразу несколько групп опубликовали данные по *B. subtilis*, полученные методом микрочипов (Helmann et al. 2001, Petersohn et al. 2001, Price et al. 2001). Во всех трех работах изменение транскрипции генов, входящих в модулон общего стресса индуцировалось этанолом и повышением температуры.

При имевшемся массиве данных было целесообразно сравнить данные, полученные методом сравнительной геномики, с данными анализа экспрессии на микрочипах. Для этого был составлен общий список генов, встреченных хотя бы в одном из трех микрочиповых исследований (241 ген). Всего про 48 генов было показано участие в

общем стрессе во всех трех работах, 51 были индуцированы стрессовым воздействием в двух работах из трех, 142 гена были причислены к регулону общего стресса только в одном исследовании из трех. На результат этого сравнения были наложены данные, полученные в результате поиска потенциальных сайтов при помощи МПВ (Таблица 5). 104 гена из 241 встреченных хотя бы в одном из трех исследований, имели в 5'-некодирующей области потенциальные сайты связывания SigB. Из 48 генов, встреченных во всех трех экспериментальных работах 37 имели потенциальный сайт связывания регулятора в 5'-некодирующей области. 5 генов, встреченных в выборках потенциально регулируемых в четырех или более геномах, были выделены в члены регулона во всех трех работах. Таким образом, гены, экспрессия которых зависела от SigB в 2 или 3 экспериментах, с относительно большей вероятностью имеют сайты, сохраненные в нескольких геномах. Более того, подавляющее большинство генов с консервативными сайтами были выделены хотя бы в одном эксперименте.

Проведенное сравнение результатов сравнительно-геномного анализа и метода микрочипов дает возможность оценить степень достоверности каждого из методов в случае изучения глобальной регуляции. Для регулона SigB *B. subtilis* определена наиболее вероятная геном-специфическая и консервативная части регулона на основе анализа трех экспериментальных работ по экспрессии на микрочипах и сравнительного анализа шести родственных *B. subtilis* геномов.

Таблица 3. Консервативное ядро регулона SigB.

	BC	BH	BA	LM	OI	SA
bmrU-bmr-bmrR	0	+	-	0	-	+
csbD (ywmG)	+	0	0	-	+	+
dps	-	+	-	+	0	(+)
gtaB	(+)	-	+	-	-	-
katB	+	+	+	-	+	0
opuE	+	+	+	0	-	+
clpP	+	+	+	-	+	-
ydaP	-	0	-	+	(+)	-
yfkM		+	-	-	+	0
ytkL	-	+	-	0	+	+
ytxG	-	+	-	+	+	+
ytxH	0	(+)	0	-	+	0
ydfO	+	-	+	0	+	0
rsbW	(+)	(+)	(+)	+	+	+
rsbVWX-sigB	0	+	0	+	+	+
yxaC	+	-	+	0	(+)	-
ycbO	(+)	-	(+)	0	0	0
yshC	+	+	+	+	-	-
yrhD	-	0	(+)	-	0	+
bioA	-	+	+	0	-	+
yvgN	-	+	-	+	+	-
mutL	-	(+)	-	(+)	+	-
yfnI	0	0	+	-	0	+

sodA	+	+	+	+	-	-
-------------	---	---	---	---	---	---

Обозначения:

"+" – присутствие потенциального σ^{32} промотора;

"-" - отсутствие потенциального σ^{32} промотора;

"0" - отсутствие гена в секвенированной части генома;

"(+)" - ген расположен в потенциально регулируемом опероне.

BC - *Bacillus cereus*, BA - *Bacillus anthracis*, BH - *Bacillus halodurans*, OI - *Oceanobacillus ihimensis*, LM - *Listeria monocytogenes*, SA - *Staphylococcus aureus*

Таблица 5. Пересечение результатов, полученных методом анализа экспрессии на микрочипах, и методом сравнительной геномики.

перед геном есть потенциальный сайт в:	Эксперсия гена зависела от SigB в:			
	в 0 исследованиях	одном исследовании	двух исследованиях	трех исследованиях
0 геномов	0	58	14	9
1 геноме	2	12	10	8
2 геномах	0	9	4	7
3 геномах	0	3	0	3
4 геномах	0	3	1	1
5 геномах	0	1	0	0

Глава IV. Регулоны устойчивости к тяжелым металлам у зубактерий. Семейство MerR (HMR)

Семейство MerR состоит из регуляторов транскрипции содержащих ДНК-связывающий домен вида спираль-поворот-спираль. Кроме металл-связывающих регуляторов, в семейство входят регулятор ответа на окислительный стресс SoxR и регуляторы множественной лекарственной устойчивости. Метал-чувствующие регуляторы выделяются внутри семейства присутствием консервативных цистеинов в позициях (Brown, 2003).

Регулоны семейства MerR чаще всего состоят из двух генов - транспортера ионов тяжелых металлов и собственно регулятора. Исключение составляют системы устойчивости к ионам ртути Hg(II) и свинца Pb (II), которые включают также ряд генов, отвечающих за изменение степени окисления иона (Hobman et al., 2003). Гены могут как располагаться в одном локусе - опероне или дивергоне, так и находиться в разных частях генома. Характерной особенностью семейства MerR является способ активации транскрипции регулируемых генов. Промотор регулируемого гена обладает спейсером длины 19 (MerR протеобактерий, CueR, HmrR, PbrR) или 20 (MerR Грам-положительных бактерий, ZntR), и следовательно, субоптимален (O'Halloran et al., 2000). Для инициации транскрипции с промоторов этого типа необходимо, чтобы белок-регулятор связался с палиндромной последовательностью, находящейся внутри промоторного спейсера (O'Halloran et al., 1999). Консенсусы палиндромных сайтов связывания специфичны для разных членов подсемейства (Brown, 2003, Hobman et al., 2003). При наличии

ионов металла-индуктора в среде регуляторный белок приобретает способность изгибать ДНК (Summers, 1992) и, таким образом, расстояние между боксами промотора становится оптимальным.

Анализ регулонов семейства MerR

Поскольку регулоны семейства MerR очень невелики, при поиске потенциальных сайтов связывания регулятора было важно учесть сложную структуру этих сайтов и соблюдать оба условия: наличие соответствующего палиндрома и промотора с заданным субоптимальным спейсером. МПВ для палиндромов были построены с использованием сайтов, взятых из (Brown, 2003), матрица для промотора была построена с использованием выборки промоторов из базы DPInteract.

При помощи программы BLAST были найдены и описаны все секвенированные на момент исследования локусы, содержащие регуляторы семейства MerR. Из них были отобраны те, у которых были обнаружены консервативные цистеины в трех позициях, ответственных за связывание с ионом металла. Было построено филогенетическое дерево регуляторов. Потенциальная специфичность регулятора определялась, исходя из его позиции на дереве и наличия в его 5'-некодирующей области соответствующего специфичного палиндромного сайта внутри промоторного спейсера. В локусах, содержащих потенциальные регуляторы, был проведен поиск комбинированного сайта промотор+палиндром, и собраны все такие случаи. В результате была предсказана специфичность нескольких транспортеров и сделано предположение о возможности вхождения в регулон CueR генов, кодирующих цитохромы c553 или c554 в *V. vulnificus* (VV20052), *V. parahaemolyticus* (VPA1012) и *V. cholerae* (VCA0766).

По результатам анализа филогенетического дерева и описанных сайтов связывания был сделан вывод об ортологичности регуляторов в парах CueR-HmrR и CadR-PbrR.

Для геномов *Providencia rettgeri*, *Pseudomonas sp.* (Q9F3U8) и *A. haloplanktis* было сделано наблюдение о смешанном типе сайта MerR - перед оперонами *mer* были найдены комбинированные сайты с палиндромом протеобактериального типа и промоторным спейсером Грам-положительного типа.

Рисунок 2. Дерево регуляторов семейства MerR. На ветках дерева обозначены специфичность регуляторов (Hg, Cu, Cd, Pb, Zn) и лого сигналов связывания. "G+" и "G-" означают Грам-положительные и Грам-отрицательные бактерии.

Выводы

1. Описана структура и состав регулонов SOS-ответа, теплового шока, общего стресса и устойчивости к тяжелым металлам в геномах рассмотренных групп.
2. Предсказана регуляция нескольких генов (*ybbN* - регулон RpoH, VV20052, VPA1012 и VCA0766 - регулон CueR).
3. Предсказана специфичность нескольких регуляторов и транспортеров, относящихся к системе устойчивости к ионам тяжелых металлов. Из 503 членов суперсемейства COG0789 для 109 предположена регуляция устойчивости к тяжелым металлам, из них 14 отнесено к подсемейству MerR Грам-положительного типа, 30 - к подсемейству MerR протеобактериального типа, 14 - к подсемейству CueR/HmrR, 6 - к подсемейству CadR/PbrR и 6 - к подсемейству ZntR. Во всех случаях определения конкретного подсемейства определены остальные потенциальные участники регулона.
4. Проведено сравнение результатов сравнительно-геномного анализа и метода микрочипов. Для регулона SigB *B. subtilis* определена наиболее вероятная геном-специфическая и консервативная части регулона на основе анализа трех экспериментальных работ по экспрессии на микрочипах и сравнительного анализа шести родственных *B. subtilis* геномов.
5. Описаны регуляторные каскады и схемы пересечения регулонов, отвечающих за тепловой стресс: регуляция гена *hrcA* сигма-фактором RpoH у *Achromobacter xylosoxidans*, *Ralstonia eutropha*, *R. solanocearum*, *Bordetella pertussis*, *B. parapertussis*, *Burkholderia pseudomallei*, *Methylobacillus flagellatus*, *Nitrosomonas eutropha*, и *Methylovorus* sp. SS1, и репрессором CtsR у *Lactococcus lactis*, *Listeria innocua*, *L. monocytogenes*, *Enterococcus faecalis*, регуляция гена *rpoH* репрессором HrcA у (*Bordetella pertussis*, *B. parapertussis* и *Methylobacillus flagellatus*), двойная регуляция генов *dnaKJ* регуляторами SigB и CtsR, и оперона *groESL* регуляторами HrcA и CtsR у *Streptococcus pyogenes*, *Staphylococcus aureus*.

Список работ, опубликованных по теме диссертации

Статьи:

1. E. Permina, A. Mironov, M. Gelfand. Damage-repair error-prone polymerases of eubacteria: association with mobile genome elements. // Gene, 2002, 293, 133–140.
2. E. Permina, M. Gelfand. Heat Shock (Sigma32 and HrcA/CIRCE) Regulons in beta-, gamma- and epsilon-Proteobacteria. // J Mol Microbiol Biotechnol. 2003; 6(3-4):174-81.

Тезисы конференции:

1. E. Permina, M. Gelfand. Comparative analysis of regulatory interactions in bacterial genomes. // Abstracts of BGRS'2002 (International Conference on Bioinformatics of Genome Regulation and Structure, Новосибирск)
2. E. Permina, M. Gelfand. Heat shock regulons in eubacteria. // Abstracts of MCCMB'03 (Moscow Conference in Computational Molecular Biology, Москва 2003)
3. A. Kazakov, O. Kalinina, E. Permina, M. S. Gelfand. Bacterial metal resistance systems regulated by transcriptional regulators of the MerR family. // Abstracts of BGRS'04.
4. A. Kazakov, E. Permina, O. Kalinina, M. S. Gelfand. // Heavy metal resistance in eubacteria: COG0789 (MerR)-family, Abstracts of MCCMB'05.
5. E. Permina. General stress regulon in Bacillus group. // Abstracts of MCCMB'05.
6. L. Sychova, E. Permina. SOS-response taxon-specific regulation of Gamma-proteobacteria. // Abstracts of MCCMB'05.